HHS Public Access

Author manuscript

Schizophr Res. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Schizophr Res. 2015 December ; 169(0): 169–177. doi:10.1016/j.schres.2015.09.008.

Severity of Thought Disorder Predicts Psychosis in Persons at Clinical High-Risk

Diana O. Perkins, MD MPH[a], **Clark D. Jeffries, PhD**[b], **Barbara A. Cornblatt, PhD**[c], **Scott W. Woods, MD**[d], **Jean Addington, PhD**[e], **Carrie E. Bearden, PhD**[f], **Kristin S. Cadenhead, MD**[g], **Tyrone D. Cannon, PhD**[d,h], **Robert Heinssen, PhD**[i], **Daniel H. Mathalon, PhD MD**[j], **Larry J. Seidman, PhD**[k], **Ming T. Tsuang, MD PhD**[g], **Elaine F. Walker, PhD**[l], and **Thomas H. McGlashan, MD**[d]

[a]Department of Psychiatry, University of North Carolina, Chapel Hill

[b]Renaissance Computing Institute, University of North Carolina, Chapel Hill NC

[c]Department of Psychiatry, Zucker Hillside Hospital, Long Island NY

[d]Department of Psychiatry, Yale University, New Haven CT

[e]Hotchkiss Brain Institute, Department of Psychiatry, University of Calgary, Alberta, Canada

[f]Departments of Psychiatry and Biobehavioral Sciences and Psychology, UCLA, Los Angeles CA

[g]Department of Psychiatry, UCSD, San Diego CA

[h]Department of Psychology, Yale University, New Haven CT

[i]National Institute of Mental Health

[j]Department of Psychiatry, UCSF, San Francisco CA

[k]Department of Psychiatry, Harvard Medical School at Beth Israel Deaconess Medical Center and Massachusetts General Hospital, Boston MA

[l]Departments of Psychology and Psychiatry, Emory University, Atlanta GA

Abstract

Background—Improving predictive accuracy is of paramount importance for early detection and prevention of psychosis. We sought a symptom severity classifier that would improve psychosis risk prediction.

Contributors: Drs. Perkins and Jeffries contributed equally to this work. Dr. Perkins and Jeffries undertook the statistical analysis, and Dr. Perkins wrote the first draft of the manuscript. All of the authors were involved in study design, contributed to and have approved the final manuscript.

Conflict of Interest: The authors declare that they have no actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three (3) years of beginning the work submitted that could inappropriately influence, or be perceived to influence, their work

Methods—Subjects were from two cohorts of the North American Prodrome Longitudinal Study. All subjects met Criteria of Psychosis-Risk States. In Cohort-1 (n=296) we developed a classifier that included those items of the Scale of Psychosis-Risk Symptoms that best distinguished subjects who converted to psychosis from nonconverters, with performance initially validated by randomization tests in Cohort-1. Cohort-2 (n=592) served as an independent test set.

Results—We derived 2-Item and 4-Item subscales. Both included unusual thought content and suspiciousness; the latter added reduced ideational richness and difficulties with focus/concentration. The Concordance Index (C-Index), a measure of discrimination, was similar for each subscale across cohorts (4-Item subscale Cohort-2: 0.71, 95%CI=[0.64,0.77], Cohort-1: 0.74, 95%CI=[0.69,0.80]; 2-Item subscale Cohort-2: 0.68, 95%CI=[0.3,0.76], Cohort-1: 0.72, 95%CI=[0.66-0.79]). The 4-Item performed better than the 2-Item subscale in 742/1000 random selections of 80% subsets of Cohort-2 subjects (p-value=1.3E-55). Subscale calibration between cohorts was proportional (higher scores/lower survival), but absolute conversion risk predicted from Cohort-1 was higher than that observed in Cohort-2, reflecting the cohorts' differences in 2-year conversion rates (Cohort-2: 0.16, 95%CI=[0.13,0.19]; Cohort-1: 0.30, 95%CI=[0.24,0.36]).

Conclusion—Severity of unusual thought content, suspiciousness, reduced ideational richness, and difficulty with focus/concentration informed psychosis risk prediction. Scales based on these symptoms may have utility in research and, assuming further validation, eventual clinical applications.

Keywords

psychosis; high-risk; risk prediction; symptom severity; schizophrenia; survival

Introduction

Development of preventative interventions for schizophrenia requires identifying persons at very high risk. An early study examining psychosis conversion in persons meeting high-risk diagnostic criteria reported a 45% 2-year conversion rate (Yung et al., 2004), however subsequent studies found 2-year conversion rates that ranged from 15-30% (Demjaha et al., 2012; DeVylder et al., 2014; Katsura et al., 2014; Lee et al., 2014; Liu et al.; Nelson et al., 2013; Riecher-Rossler et al., 2009; Ruhrmann et al., 2010; Woods et al., 2009; Ziermans et al., 2011). Efforts are needed to improve psychosis risk prediction.

Prominent among the scales used to evaluate symptoms associated with psychosis risk is the Scale of Psychosis-Risk Symptoms (SOPS) (McGlashan et al., 2010; Miller et al., 2002). The SOPS comprises 19 symptoms in four domains that include: *positive* (unusual thought content/delusional ideas, suspiciousness/persecutory ideas, grandiose ideas, perceptual abnormalities/hallucinations, disorganized communication), *negative* (social anhedonia, avolition, decreased expression of emotion, decreased experience of emotions and self, reduced ideational richness, reduced occupational functioning), *disorganized* (odd behavior or appearance, bizarre thinking, trouble with focus and attention, impaired hygiene), and *general* (sleep disturbance, dysphoric mood, motor disturbances, impaired stress tolerance). The symptoms evaluated by the SOPS were chosen to reflect broadly the symptoms experienced by persons with schizophrenia during their prodrome.

We sought to identify among items measured by the SOPS subsets that best predicted psychosis conversion. We considered two large independent cohorts, the North American Prodrome Longitudinal Study Cohort-1 and the North American Prodrome Longitudinal Study Cohort-2. This allowed construction of risk prediction subscales in Cohort-1 and evaluation of subscale performance in Cohort-2.

Methods

1_Subjects

Detailed study methods were reported previously (Addington et al., 2007; Addington et al., 2012; Cannon et al., 2008). In brief, the North American Prodrome Longitudinal Study is a multisite observational study of the predictors and mechanisms of conversion to psychosis in persons meeting Criteria of Psychosis-Risk Syndromes (COPS) (Miller et al., 2003). There were two non-overlapping waves of recruitment, Cohort-1 and Cohort-2. For Cohort-1 a database combined the results *post hoc* from eight independent studies that used a prospective design and similar ascertainment and rating methods (Addington et al., 2007). Cohort-2 was developed as a 2-year prospective collaboration of the same eight sites (Addington et al., 2012). For both cohorts, subjects' ages ranged from 12 to 35. Studies were approved by the sites' Institutional Review Boards, and subjects provided written informed consent or assent, with a parent/guardian consenting for persons under age 18.

Study participants were evaluated using the Structured Interview for Psychosis-Risk Syndromes (SIPS) (McGlashan et al., 2010; Miller et al., 2002) to determine if they met criteria for one or more of the following high-risk syndromes: attenuated psychotic symptoms syndrome; brief intermittent psychotic symptoms syndrome; and genetic risk and deterioration syndrome. The Presence of Psychosis (POP) criteria (McGlashan et al., 2010; Miller et al., 2002) were used to classify a subject as a "converter" to psychosis (see Supplement for detailed criteria). For subjects who converted, date of conversion was estimated by clinical interview and, if available, medical records. *Diagnostic and Statistical Manual IV* (First, 2002) (DSM-IV) psychotic disorder diagnosis was based on Structured Clinical Interview for DSM IV (First, 2002) performed by trained raters. Subjects were re-assessed every six months by raters. While symptom severity was assessed at baseline, prior to conversion, study raters had access to baseline ratings when evaluating conversion status. There is a possibility that this knowledge could have impacted assessment of conversion. To protect against possible bias all high-risk subjects were reviewed at study entry and at conversion by experts (JA and TM) during a diagnostic conference call, to ensure that criteria were met.

The severity of symptoms was scored on the Scale of Psychosis-risk Symptoms (SOPS) (McGlashan et al., 2010) as follows: 0 =*absent*; 1=*questionably present*; 2=*mild*; 3=*moderate*; 4=*moderately severe*; 5=*severe but not psychotic*; and 6=*severe and psychotic*. To simplify analysis, we rescored mild and questionably present from "1" or "2" to "0". We rescored the threshold severity as follows: "moderate" as "1"; "moderately severe" as "2"; "severe but not psychotic" as "3"; and "severe and psychotic" as "4". Applying our analysis strategy to the original scale had no effect on choice of informative symptoms.

For Cohort-1, raters at each site were trained by the instrument's developers and achieved high inter-rater reliability for high-risk syndrome diagnoses (kappa >0.80) (Addington et al., 2007; Cannon et al., 2008). In addition several sites participated in an evaluation of symptom scoring reliability, achieving intra-class coefficients of > 0.7 for each item (Miller et al., 2003). For Cohort-2, raters were required to have yearly assessments; intraclass correlation coefficients for the SOPS total and positive subscales were required to be >0.8 (Addington et al., 2012).

We excluded from this report subjects who did not meet Criteria of Psychosis-Risk States, who had no follow-up visits, or who had items missing from the baseline SOPS (Figure 1). The follow-up period for survival analysis was two years (the duration of systematic follow-up for Cohort-2).

2_Statistical Methods

2.1 Classifier Development—We sought a "risk prediction subscale" for the SOPS, meaning a sum of chosen items that best identified high-risk subjects who subsequently developed psychosis. We used a simple "greedy algorithm" (Comen et al., 2009; Liu et al., 2005) that first finds the best single item relative to a specified metric. Then, if possible, it finds a second item that, when added to the first, most improves the metric, and so on. The algorithm terminates when no additional items improve the metric. Classifier development implemented the greedy algorithm using five-fold cross validation (Kohavi, 1995) with Excel macros and add-ins (Moons et al., 2012). We excluded a random 25% of the subjects from each group, then randomly partitioned the remaining subjects each into five nearly equal subsets. Four converter and four nonconverter subsets were selected and the algorithm applied to all of the 25 possible combinations of converter and nonconverter subsets. We then randomly re-partitioned the five subsets and repeated the symptom selection process, a total of 20 times, resulting in 500 trials. We then excluded a new random 25% of subjects, and then repeated the entire process 10 times, thus generating 5000 total trials. As each model was built *ab initio* from subsets of the data, the derived classifiers were not identical. There is a wealth of literature on the merits of various model-building strategies (Hand, 2006; Harrell et al., 1984; Harrell et al., 1996; Kohavi, 1995) but less guidance on strategies to best integrate the multiple derived classifiers. Our approach was to rank the symptoms by their selection frequencies, with the most frequently selected items forming the integrated classifiers.

As part of the classifier development phase (Cohort-1) we used a randomization test (Fisher, (1971) [1935]) to determine whether the derived subscales actually performed better than chance. We did so because modern algorithms are capable of finding patterns even in randomized data due to hidden interrelationships. The area under the curve (AUC) of the receiver operating characteristic (ROC) is a plot of sensitivity (predicted positives/true positives) and 1-specificity (predicted negatives/true negatives), at each possible cut-off point for the scale score. From samples of real data, the typical AUC can be in excess of 0.5, although 0.5 is the expected null value from random classification using prior probability (Rucker et al., 2007). A randomization test requires that pseudo-classifiers are constructed *ab initio* from pseudo-data with exactly the same algorithm used for true data (Buzkova et

al., 2011; Lindgren et al., 1996; Rucker et al., 2007; Smit et al., 2008; Tropsha, 2010). Applying this process 1000 times to Cohort-1 data, we created pseudo-data by randomly assigning subjects to pseudo-groups of "converted" or "nonconverted," preserving original group sizes. Exactly the same classifier construction process as above was applied to the pseudo-data to yield 1000 pseudo-classifiers.

2.2 Survival Analysis—Validation was done with survival analyses using R version 3.1.2. We used two related measures to evaluate discrimination (Heagerty and Zheng, 2005). To evaluate the ability of the prediction model to order the survival time we used the Concordance Index (C-index) (Pencina and D'Agostino, 2004); and to order the survival status (converter/nonconverter) the AUC (Blanche et al., 2013). Both the C-Index and the AUC range from 0.5 (no predictive ability)-1.0 (perfect predictive ability). The success rate difference (SRD) is the difference in conversion rates for subjects at high and low risk as determined at a specified cut-off value for the scale, thus ranging from 0-1 (Kraemer and Kupfer, 2006). As a measure of utility we used the Number Needed to Take (NNT) (analagous to the more familiar "Number Needed to Treat", a utility measure often used in clinical trials); note that NNT=1/SRD (Kraemer and Kupfer, 2006). The NNT indicates how many persons need to be identified as high-risk to detect one conversion more than that seen in the low-risk group. The calibration plot, a scatterplot of the Kaplan-Meier survival estimates at different classifier cutoff points, compares scaling of the classifiers in the two cohorts.

Results

3.1 Study subjects (Table 1)

Baseline evaluations for Cohort-1 occurred in 1998-2005 and for Cohort-2 in 2008-2013. Compared to included subjects, excluded subjects had significantly lower parental education in both cohorts. For Cohort-1, the diagnosis at conversion was known for 59 (66%): Bipolar Disorder (n=6), Brief Psychotic Disorder (n=2), Delusional Disorder (n=2), Psychosis NOS (n=16), Schizoaffective Disorder (n=6), Schizophrenia (n=15), and Schizophreniform Disorder (n=12). For Cohort-2, the diagnosis at conversion was known for 78 (85%): Bipolar Disorder (n=7), Brief Psychotic Disorder (n=2), Delusional Disorder (n=3), Psychosis NOS (n=31), Schizoaffective Disorder (n=5), Schizophrenia (n=18), and Schizophreniform Disorder (n=12).

3.2 Classifier development

The greedy algorithm was run repeatedly on randomly chosen subsets of converters and nonconverters. With each run, somewhat different combinations of symptoms were chosen (Figure 2). We observed greedy algorithm termination after choosing an average of 5.9 symptoms (of 19) (sd = 0.85). Suspiciousness/persecutory ideas (P2) and unusual thought content (P1) and were chosen in almost all trials followed by reduced ideational richness (N5) and trouble with focus and attention (D3). (See Supplement for detailed descriptions of these symptoms.) Considering Figure 2, we defined a 2-Item subscale classifier as the sum of the two most frequently selected symptoms (P1,P2) and a 4-Item subscale as the sum of

the four most frequently selected symptoms (P1,P2,N5,D3). The same four were also first selected when the greedy algorithm was applied to the full data.

3.3 Validation of Classifier Development Methods in Cohort-1

Computer-implemented classifier methods can find seemingly robust patterns in random data, but the AUCs of random data should be low compared to the AUC of a classifier built with true data. We applied exactly the same greedy algorithm to data where converter/nonconverter group membership was randomly reassigned and calculated the resulting AUC. The AUCs of each pseudo-classifier applied to its pseudo-data were summarized in a histogram and fitted with a beta distribution (Figure 3). The AUC of the true 4-Item subscale applied to Cohort-1 was 0.74 with a parametric p-value relative to the distribution of 3.9E-5. Alternatively, since in exactly one of the 1000 trials, the pseudo-classifier AUC was by chance better than the true AUC, a nonparametric p-value is (1+1)/(1000+1)=2.0E-3.

3.4 Independent testing of classifier performance in Cohort-2

Cohort-2 baseline conversion rates at all time-points were less than those of Cohort-1 (Cannon et al., 2008; Woods et al., 2009)] (Figure 4).

In Cohort-2 discrimination, as evaluated with the C-Index, was consistently higher with the 4-Item than with the 2-Item subscale at all follow-up time points (Supplement, Figure S1). At 2-years the C-Index for the 4-Item subscale was greater than the 2-Item subscale for 743 of 1000 trials of 80% (with replacement) randomly chosen subsets of subjects (p-value=1.3E-55). In Cohort-2 at 2-years the 4-Item subscale was greater than the SOPS Total in 968 of 1000 trials, and the 2-Item subscale greater than the SOPS Total in 883 of 1000 trials (p-values<1.3E-55).

Using an optimized cut-off, the conversion rates over time were significantly higher for persons declared “high-risk” compared to persons compared “low-risk” at times 6, 12, 18 and 24 months (Figure 4). The AUCs did not vary substantially over time.

Calibration curves compared scaling with a graph of the Kaplan-Meier survival (herein nonconversion) proportions at 6, 12, 18, and 24 months, with values from Cohort-1 (predicted) on the x-axis and Cohort-2 (observed) on the y-axis (Figure 5). The Pearson R^2 values, reported in the Figure legends for each time period, were all above 0.9, indicating that higher subscale scores predicted lower survival proportionally in Cohort-1 and Cohort-2. Miscalibration-in-the-large was apparent, however, related to the higher survival rates seen in Cohort-2 relative to Cohort-1.

Discussion

The performance of the 2-Item subscale indicates that the severity of unusual thought content (P1), referential thinking (both P1,P2) and suspiciousness (P2) are key high-risk symptoms. The majority of published studies examining symptoms and risk prediction likewise have reported that items reflecting disordered thought content (unusual ideas (Katsura et al., 2014; Nelson et al., 2013; Salokangas et al., 2013; Thompson et al., 2013; Thompson et al., 2011; Wilcox et al., 2014), suspiciousness (Riecher-Rossler et al., 2009;

Salokangas et al., 2013; Wilcox et al., 2014), bizarre thoughts (Ruhrmann et al., 2010), odd beliefs/magical thinking (Mason et al., 2004), problems distinguishing fantasy and reality (Klosterkotter et al., 2001), unstable ideas of reference (Klosterkotter et al., 2001), derealization (Klosterkotter et al., 2001)) are more severe in converters than nonconverters; this is despite variations in: diagnostic criteria for clinical high-risk; instruments to measure symptom severity; and study populations (Schultze-Lutter et al., 2013) (Supplement, Table S1). It is important to note that our reported performance of the subscales applied to persons meeting Criteria of Psychosis-Risk States, and may be different in persons meeting other high-risk diagnostic criteria.

Discrimination with the 4-Item subscale was, with certainty, better than that of the 2-Item subscale; however, the magnitude of this difference was small, implying clinical importance is yet to be determined. The 4-Item subscale (and the 2-Item subscale) also performed better than the SOPS Total. However, we do not claim to have found optimal or unique combinations of symptoms for psychosis risk prediction. Furthermore, there may be high-risk subtypes better predicted by different symptoms.

There is further reason to continue to consider the two additional items in future studies. The additional items reflect disturbances of thought process. N5 (reduced ideational richness) is defined in the SOPS as having difficulties in: following everyday conversations; making sense of familiar phrases; grasping the gist of conversations; escaping patterns of repetitious or simplistic thought content; considering alternative positions; shifting ideas; using anything but simplistic language; and thinking abstractly. D3 (trouble with focus and attention) includes difficulties in: maintaining focused attention and resisting distraction due to internal and external stimuli; holding conversations in memory; and executing other short-term memory tasks. Other investigators have predicted psychosis risk from disordered thought processes including: poor attention (Yung et al., 2004); disorganized cognitive subscale (Demjaha et al., 2012); interference, perseveration, blockage, and pressure of thoughts; disturbances of receptive language (Klosterkotter et al., 2001); and conceptual disorganization (Nelson et al., 2013). Supporting the value of including disturbances in both thought content and thought processes in defining the high-risk state is a recent study (Schultze-Lutter et al., 2014) integrating the Criteria of Psychosis-Risk Syndromes (used in this study) and COGDIS criteria, namely, meeting criteria for at least two of nine symptoms from a list that includes “unstable ideas of reference” (partially overlapping with P1 and P2) and eight symptoms reflecting disturbances of thought processes. Notably, the COGDIS items “thought Interference” and “disturbance of receptive speech” partially overlap with D3, and items “disturbances of abstract thinking” and “disturbance of receptive speech” partially overlap with N5.

It is notable that some symptoms were seldom or never selected (Figure 2). In particular, P4 (perceptual disturbances) is absent from our risk prediction subscales as it was infrequently selected by the greedy algorithm. With two exceptions (Klosterkotter et al., 2001; Mason et al., 2004), most other studies have likewise failed to find perceptual disturbances as predictive (DeVylder et al., 2014; Katsura et al., 2014; Nelson et al., 2013; Riecher-Rossler et al., 2009; Ruhrmann et al., 2010; Salokangas et al., 2013; Thompson et al., 2013; Velthorst et al., 2009; Wilcox et al., 2014), including a factor-analysis where a factor

including perceptual abnormalities, mood swings/lability, aggression/dangerous behaviors, and suicidality/self-harm (symptoms common in personality disorders) was not predictive of one-year conversion (Raballo et al., 2011). Psychosis high-risk diagnostic criteria (Addington, 2004; Carpenter and Tandon, 2013), including those used in this study, would classify a person with perceptual abnormalities alone as high-risk. These findings raise the question of whether perceptual disturbances alone should be a diagnostic criteria for a clinical high-risk syndrome.

As shown in Figure 3, we found that many models built using pseudo-data had AUCs above 0.6. Datasets with hidden relationships may show AUCs well above the hypothetical null result 0.5 (Rucker et al., 2007). Randomization testing is a way to prove—to a certain p-value— that the classifier performance in the test set is not unlikely by chance, thus facilitating classifier development (Rucker et al., 2007; Tropsha, 2010).

The 2-year conversion in Cohort-2 (16%) is at the low end of rates reported in recent studies that range from 15-26% (Demjaha et al., 2012; DeVylder et al., 2014; Katsura et al., 2014; Nelson et al., 2013; Ziermans et al., 2011). The calibration slope was similar in both cohorts, meaning that, regardless of overall conversion rates, higher scores indicate proportionally greater increase in psychosis risk. However calibration-in-the-large was dissimilar between cohorts, reflecting in part the higher conversion rate in Cohort-1 than in Cohort-2. Case identification strategies may influence the case-mix relative to psychosis risk; for example, case identification through screenings of clinic or general populations may yield a broader case-mix with lower conversion rates, possibly impacting calibration-in-the-large. The emerging risk algorithms, biological or clinical, need to consider calibration before they are used to provide absolute estimates of psychosis risk.

In contemporary research settings, the 2-Item and 4-Item subscales might have utility in identifying higher-risk subgroups in persons meeting Criteria of Psychosis-Risk Syndromes. Considering Cohort-2; with the cutoffs presented in Figure 4d about a third of subjects met the severity threshold with a 2-year conversion risk of 30% compared to 9% in the two-third subjects identified as at lower risk. At 2-year follow-up, in Cohort-2 the NNT for the 2-Item and 4-Item subscales were 5.3 and 4.8, respectively (corresponding to a medium effect size), both lower than that of the SOPS Total (NNT=12.5), (corresponding to small effect size) (Kraemer and Kupfer, 2006).

Use of risk prediction subscales in clinical settings will require development of strategies to educate community mental health care providers about assessment of high-risk symptoms and diagnosis of a clinical high-risk state. In addition, discrimination and calibration of the risk prediction subscales would need to be evaluated in the hands of trained community providers prior to any recommendations about clinical usefulness (Salokangas et al., 2013). In particular, calibration and discrimination may differ in community settings. About half of persons diagnosed with psychotic disorders have sought mental health care prior to onset of psychosis (Rietdijk et al., 2011), and as many as 4-8% of adolescents and young adults seeking mental health care may meet clinical high-risk criteria (Ising et al., 2012; Rietdijk et al., 2014); it is unclear whether these patient pools are fully represented in psychosis-risk research. However, the potential value of symptom-based risk prediction to clinical practice

is clear. That potential may be realized when scoring systems are developed that consider the personal, social, and financial benefits of treatment (e.g. likelihood of psychosis prevention) as well as costs (direct medical costs, side-effects, etc.) (Essock et al., 2002; McNeil and Kaij, 1979). Applications of the present work might include treatment monitoring, integration with other evaluations, and programs of stepwise application of treatments, all in the context of prudent counseling (McGorry et al., 2009).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

U01 MH081902 (Cannon), P50 MH066286 (Bearden), U01 MH081857 (Cornblatt), U01 MH82022 (Woods), U01 MH066134 (Addington), U01 MH081944 (Cadenhead), R01, U01 MH066069 (Perkins, Jeffries), MH076989 (Mathalon), U01 MH081928 (Seidman), U01 MH081988 (Walker).

Role of the Funding Source. This project was a cooperative agreement between the investigator sites and the National Institutes of Health.

References

Addington J. The diagnosis and assessment of individuals prodromal for schizophrenic psychosis. CNS spectrums. 2004; 9(8):588–594. [PubMed: 15273651]

Addington J, Cadenhead KS, Cannon TD, Cornblatt B, McGlashan TH, Perkins DO, Seidman LJ, Tsuang M, Walker EF, Woods SW, Heinssen R. North American Prodrome Longitudinal Study: a collaborative multisite approach to prodromal schizophrenia research. Schizophrenia bulletin. 2007; 33(3):665–672. [PubMed: 17255119]

Addington J, Cadenhead KS, Cornblatt BA, Mathalon DH, McGlashan TH, Perkins DO, Seidman LJ, Tsuang MT, Walker EF, Woods SW, Addington JA, Cannon TD. North American Prodrome Longitudinal Study (NAPLS 2): Overview and recruitment. Schizophrenia research. 2012; 142(1-3): 77–82. [PubMed: 23043872]

Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Statistics in medicine. 2013; 32(30):5381–5397. [PubMed: 24027076]

Buzkova PL, Lumley T, Rice K. Permutation and parametric bootstrap tests for gene-gene and gene-environment interactions. Ann Hum Genet. 2011; 75(1):36–45. [PubMed: 20384625]

Cannon TD, Cadenhead K, Cornblatt B, Woods SW, Addington J, Walker E, Seidman LJ, Perkins D, Tsuang M, McGlashan T, Heinssen R. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. Archives of general psychiatry. 2008; 65(1):28–37. [PubMed: 18180426]

Carpenter WT, Tandon R. Psychotic disorders in DSM-5: Summary of changes. Asian journal of psychiatry. 2013; 6(3):266–268. [PubMed: 23642992]

Comen, TH.; Leiserson, CE.; Rivest, RL.; Stein, C. Greedy Algorithms, Introduction to Algorithms. Massachusetts Institute of Technology; Cambridge, Massachusetts: 2009. p. 414-450.

Demjaha A, Valmaggia L, Stahl D, Byrne M, McGuire P. Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. Schizophrenia bulletin. 2012; 38(2):351–359. [PubMed: 20705805]

DeVylder JE, Muchomba FM, Gill KE, Ben-David S, Walder DJ, Malaspina D, Corcoran CM. Symptom trajectories and psychosis onset in a clinical high-risk cohort: the relevance of subthreshold thought disorder. Schizophrenia research. 2014; 159(2-3):278–283. [PubMed: 25242361]

Essock, SM.; Frisman, LK.; Covell, NH. The economics of the treatment of schizophrenia, Neuropsychopharmacology: The Fifth Generation of Progress. American College of Neuropsychopharmacology; 2002. p. 809-818.

First, MB.; Spitzer, RL.; Givvon, M.; Williams, JBW. Biometrics Research. New York State Psychiatric Institute; New York: 2002. Structured Clinical Interview for DSM-IV TR Axis I Disorders, Non-patient Edition (SCID-I/NP).

Fisher, RA. The Design of Experiments (9th ed.). 9. Macmillian; 1971. 1935

Hand DJ. Classifier technology and the illusion of progress. Statistical Science. 2006; 21:1–34. [PubMed: 17906740]

Harrell FE Jr. Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Statistics in medicine. 1984; 3(2):143–152. [PubMed: 6463451]

Harrell FE Jr. Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996; 15(4):361–387. [PubMed: 8668867]

Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005; 61(1): 92–105. [PubMed: 15737082]

Ising HK, Veling W, Loewy RL, Rietveld MW, Rietdijk J, Dragt S, Klaassen RM, Nieman DH, Wunderink L, Linszen DH, van der Gaag M. The validity of the 16-item version of the Prodromal Questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general help-seeking population. Schizophrenia bulletin. 2012; 38(6):1288–1296. [PubMed: 22516147]

Katsura M, Ohmuro N, Obara C, Kikuchi T, Ito F, Miyakoshi T, Matsuoka H, Matsumoto K. A naturalistic longitudinal study of at-risk mental state with a 2.4 year follow-up at a specialized clinic setting in Japan. Schizophrenia research. 2014; 158(1-3):32–38. [PubMed: 25034763]

Klosterkotter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F. Diagnosing schizophrenia in the initial prodromal phase. Archives of general psychiatry. 2001; 58(2):158–164. [PubMed: 11177117]

Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. 1995; 12(2):1137–1143.

Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biological psychiatry. 2006; 59(11):990–996. [PubMed: 16368078]

Lee TY, Kim SN, Correll CU, Byun MS, Kim E, Jang JH, Kang DH, Yun JY, Kwon JS. Symptomatic and functional remission of subjects at clinical high risk for psychosis: a 2-year naturalistic observational study. Schizophrenia research. 2014; 156(2-3):266–271. [PubMed: 24815568]

Lindgren F, Hansen B, Sjostrom M. Model validation by permutation tests. Journal of Chemometrics. 1996; 10:521–532. W., K. L., E.

Liu CC, Lai MC, Liu CM, Chiu YN, Hsieh MH, Hwang TJ, Chien YL, Chen WJ, Hua MS, Hsiung PC, Huang YC, Hwu HG. Follow-up of subjects with suspected pre-psychotic state in Taiwan. Schizophrenia research. 2011; 126(1-3):65–70. [PubMed: 21112187]

Liu X, Krishnan A, Mondry A. An entropy-based gene selection method for cancer classification using microarray data. BMC bioinformatics. 2005; 6:76. [PubMed: 15790388]

Mason O, Startup M, Halpin S, Schall U, Conrad A, Carr V. Risk factors for transition to first episode psychosis among individuals with 'at-risk mental states'. Schizophrenia research. 2004; 71(2-3): 227–237. [PubMed: 15474894]

McGlashan, TH.; Walsh, BC.; Woods, SW. The Psychosis Risk Syndrome: Handbook for Diagnosis and Follow-Up. Oxford UP; 2010.

McGorry PD, Nelson B, Amminger GP, Bechdolf A, Francey SM, Berger G, Riecher-Rossler A, Klosterkotter J, Ruhrmann S, Schultze-Lutter F, Nordentoft M, Hickie I, McGuire P, Berk M, Chen EY, Keshavan MS, Yung AR. Intervention in individuals at ultra-high risk for psychosis: a review and future directions. The Journal of clinical psychiatry. 2009; 70(9):1206–1212. [PubMed: 19573499]

McNeil TF, Kaij L. Etiological relevance of comparisons of high-risk and low-risk groups. Acta psychiatrica Scandinavica. 1979; 59(5):545–560. [PubMed: 463591]

Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, McFarlane W, Perkins DO, Pearlson GD, Woods SW. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. Schizophrenia bulletin. 2003; 29(4):703–715. [PubMed: 14989408]

Miller TJ, McGlashan TH, Rosen JL, Somjee L, Markovich PJ, Stein K, Woods SW. Prospective diagnosis of the initial prodrome for schizophrenia based on the Structured Interview for Prodromal Syndromes: preliminary evidence of interrater reliability and predictive validity. The American journal of psychiatry. 2002; 159(5):863–865. [PubMed: 11986145]

Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart (British Cardiac Society). 2012; 98(9):683–690. [PubMed: 22397945]

Nelson B, Yuen HP, Wood SJ, Lin A, Spiliotacopoulos D, Bruxner A, Broussard C, Simmons M, Foley DL, Brewer WJ, Francey SM, Amminger GP, Thompson A, McGorry PD, Yung AR. Long-term Follow-up of a Group at Ultra High Risk ("Prodromal") for Psychosis: The PACE 400 Study. JAMA psychiatry. 2013; 70(8):793–802. [PubMed: 23739772]

Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Statistics in medicine. 2004; 23(13): 2109–2123. [PubMed: 15211606]

Raballo A, Nelson B, Thompson A, Yung A. The comprehensive assessment of at-risk mental states: from mapping the onset to mapping the structure. Schizophrenia research. 2011; 127(1-3):107–114. [PubMed: 21295947]

Riecher-Rossler A, Pflueger MO, Aston J, Borgwardt SJ, Brewer WJ, Gschwandtner U, Stieglitz RD. Efficacy of using cognitive status in predicting psychosis: a 7-year follow-up. Biological psychiatry. 2009; 66(11):1023–1030. [PubMed: 19733837]

Rietdijk J, Fokkema M, Stahl D, Valmaggia L, Ising HK, Dragt S, Klaassen RM, Nieman DH, Loewy R, Cuijpers P, Delespaul P, Linszen DH, van der Gaag M. The distribution of self-reported psychotic-like experiences in non-psychotic help-seeking mental health patients in the general population; a factor mixture analysis. Social psychiatry and psychiatric epidemiology. 2014; 49(3): 349–358. [PubMed: 24126556]

Rietdijk J, Hogerzeil SJ, van Hemert AM, Cuijpers P, Linszen DH, van der Gaag M. Pathways to psychosis: help-seeking behavior in the prodromal phase. Schizophrenia research. 2011; 132(2-3): 213–219. [PubMed: 21907547]

Rucker CL, Rucker G, Meringer M. y-Randomization and its variants in QSPR/QSAR. J Chem Inf Model. 2007; 47(6):2345–2357. [PubMed: 17880194]

Ruhrmann S, Schultze-Lutter F, Salokangas RK, Heinimaa M, Linszen D, Dingemans P, Birchwood M, Patterson P, Juckel G, Heinz A, Morrison A, Lewis S, von Reventlow HG, Klosterkotter J. Prediction of psychosis in adolescents and young adults at high risk: results from the prospective European prediction of psychosis study. Archives of general psychiatry. 2010; 67(3):241–251. [PubMed: 20194824]

Salokangas RK, Dingemans P, Heinimaa M, Svirskis T, Luutonen S, Hietala J, Ruhrmann S, Juckel G, Graf von Reventlow H, Linszen D, Birchwood M, Patterson P, Schultze-Lutter F, Klosterkotter J, group E. Prediction of psychosis in clinical high-risk patients by the Schizotypal Personality Questionnaire. Results of the EPOS project. Eur Psychiatry. 2013; 28(8):469–475. [PubMed: 23394823]

Schultze-Lutter F, Klosterkotter J, Ruhrmann S. Improving the clinical prediction of psychosis by combining ultra-high risk criteria and cognitive basic symptoms. Schizophrenia research. 2014; 154(1-3):100–106. [PubMed: 24613572]

Schultze-Lutter F, Schimmelmann BG, Ruhrmann S, Michel C. 'A rose is a rose is a rose', but at-risk criteria differ. Psychopathology. 2013; 46(2):75–87. [PubMed: 22906805]

Smit S, Hoefsloot HC, Smilde AK. Statistical data processing in clinical proteomics. Journal of chromatography. B, Analytical technologies in the biomedical and life sciences. 2008; 866(1-2): 77–88.

Thompson A, Nelson B, Bruxner A, O'Connor K, Mossaheb N, Simmons MB, Yung A. Does specific psychopathology predict development of psychosis in ultra high-risk (UHR) patients? The Australian and New Zealand journal of psychiatry. 2013; 47(4):380–390. [PubMed: 23399857]

Thompson A, Nelson B, Yung A. Predictive validity of clinical variables in the "at risk" for psychosis population: international comparison with results from the North American Prodrome Longitudinal Study. Schizophrenia research. 2011; 126(1-3):51–57. [PubMed: 21035313]

Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics. 2010; 29(6-7):476–488.

Velthorst E, Nieman DH, Becker HE, van de Fliert R, Dingemans PM, Klaassen R, de Haan L, van Amelsvoort T, Linszen DH. Baseline differences in clinical symptomatology between ultra high risk subjects with and without a transition to psychosis. Schizophrenia research. 2009; 109(1-3): 60–65. [PubMed: 19272756]

Wilcox J, Briones D, Quadri S, Tsuang M. Prognostic implications of paranoia and thought disorder in new onset psychosis. Comprehensive psychiatry. 2014; 55(4):813–817. [PubMed: 24439562]

Woods SW, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Heinssen R, Perkins DO, Seidman LJ, Tsuang MT, Walker EF, McGlashan TH. Validity of the prodromal risk syndrome for first psychosis: findings from the North American Prodrome Longitudinal Study. Schizophrenia bulletin. 2009; 35(5):894–908. [PubMed: 19386578]

Yung AR, Phillips LJ, Yuen HP, McGorry PD. Risk factors for psychosis in an ultra high-risk group: psychopathology and clinical features. Schizophrenia research. 2004; 67(2-3):131–142. [PubMed: 14984872]

Ziermans TB, Schothorst PF, Sprong M, van Engeland H. Transition and remission in adolescents at ultra-high risk for psychosis. Schizophrenia research. 2011; 126(1-3):58–64. [PubMed: 21095104]

Figure 1a.

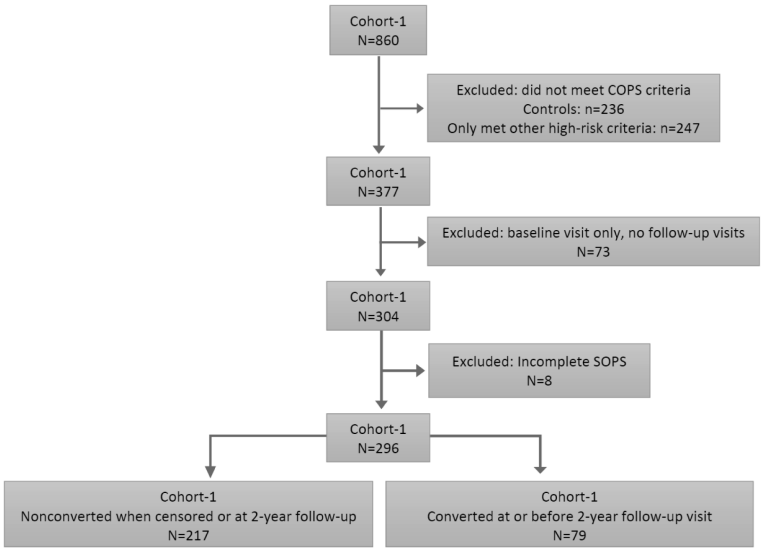

Figure 1b.

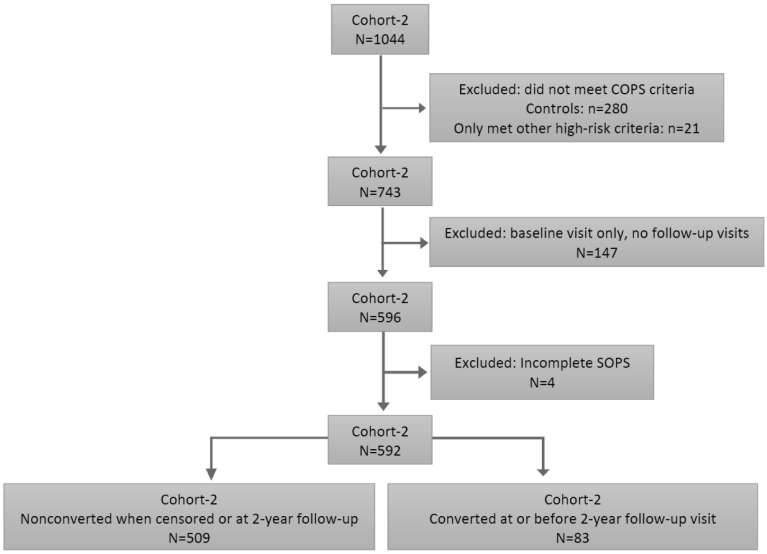

Figure 1.
CONSORT Diagram of subjects included in Cohort-1 and Cohort-2.

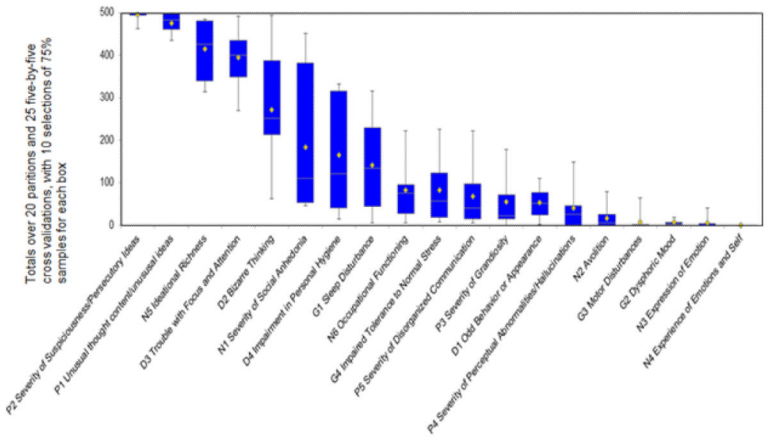

Figure 2.

Shown is the quartile plot reflecting the number of times each symptom was chosen for the subscale (maximum was 500 times per run) over the 10 runs. Notably, the symptoms P2, P1, N5, D3 dominated the choices, implying that all four are somewhat informative of transition to psychosis. Other symptoms (e.g. N4) were seldom chosen as informative.

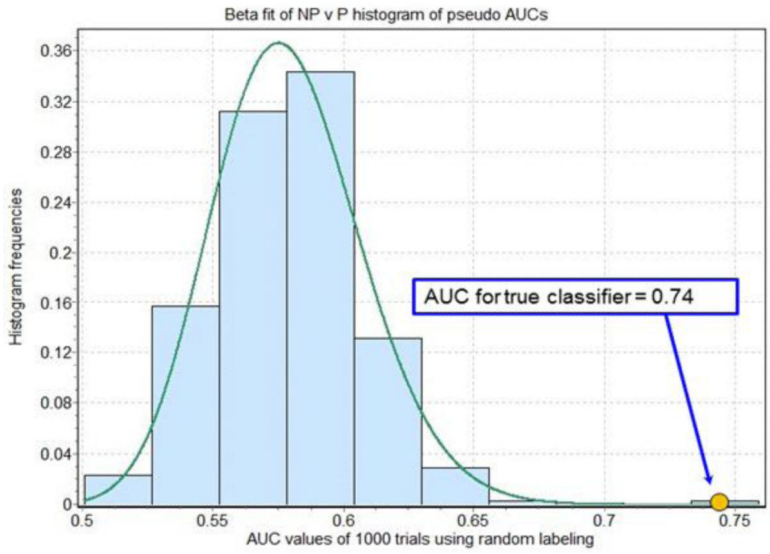

Figure 3.

Randomization test results of classifier development using data from Cohort-1 with group (converter or nonconverter) randomly re-assigned, with AUC as metric. The histogram was fitted accurately with a beta distribution (both Kolmogorov-Smirnov and Anderson-Darling $\alpha > 0.01$). The true classifier as derived with P2, P1, N5, D3 applied to the true data achieved AUC = 0.74, having distribution p-value = 3.9E-5. Only once in 1000 trials did a pseudo-classifier achieve a higher AUC, implying a nonparametric p-value = 0.002. Note that, due to hidden interdependencies in the data, the pseudo-classifiers built with random data frequently gave AUCs greater than 0.6, well above the customary and hypothetical "random" AUC of 0.5.

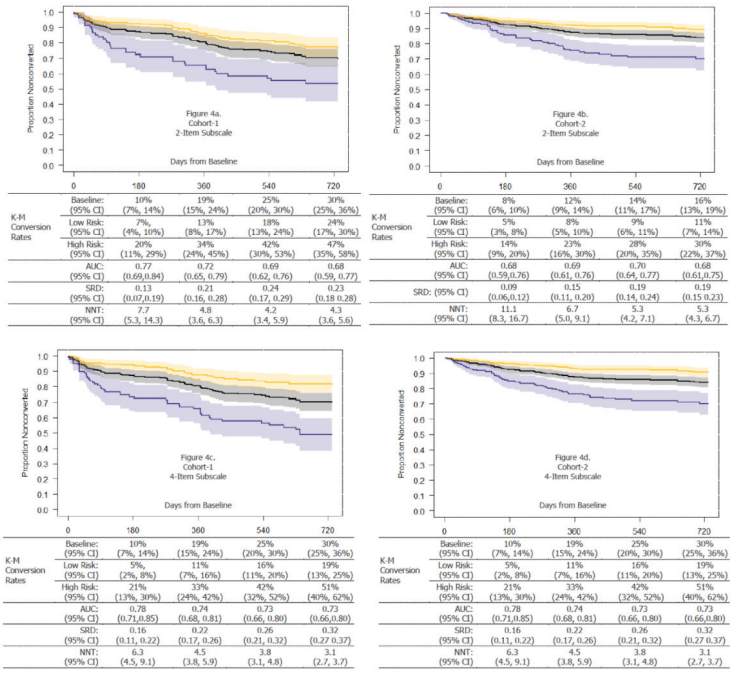

Figure 4.

Kaplan Meier Survival Curves. For the 2-Item subscale, with a cutoff of 3, there were 80 (27%) Cohort-1 and 166 (28%) Cohort-2 subjects at “High Risk”. For the 4-Item subscale, with a cutoff of 4, there were 98 (33%) Cohort-1 and 193 (33%) Cohort-2 subjects at “High Risk”. Shadowed region indicates 95% confidence intervals. AUC (Area Under the Curve of the Receiver Operating Characteristic) ranges from 0.5 to 1. The SRD (Success Rate Difference) is the difference between survival in the high-risk and low-risk groups, and ranges from 0-1. NNT (Number Needed to Take) indicates the number of persons that need to be declared as high- and low-risk for one additional true converter to be detected in the high-risk group. Black line indicates baseline, gold line indicates Low-Risk, and blue line indicates High-Risk Kaplan-Meier (K-M) survival curves.

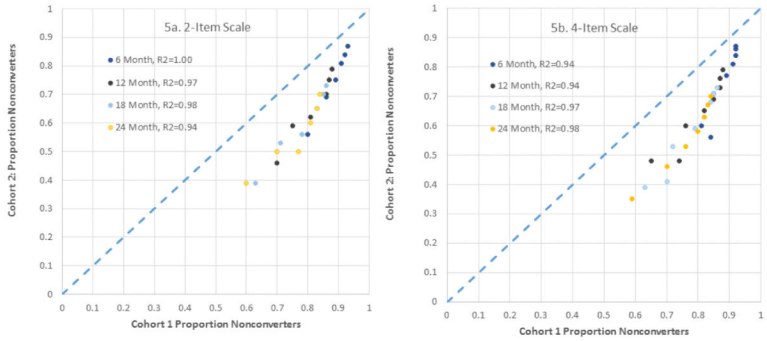

Figure 5.

Calibration curves comparing the observed performance in Cohort 2 with the predicted performance of the 2-Item and 4-Item subscales from Cohort-1. Each point represents the survival at particular time points (6, 12,18, & 24 months) at a specific cutoff point for the subscales (observed scores for the 2-Item subscale ranged from 0-≥5, for the 4-Item subscale ranged from 0-≥7). The blue diagonal line indicates perfect calibration. The Pearson correlation of values from Cohort-1 and Cohort-2 at each time point is given in the legends.

Table 1

Demographic characteristics of included and excluded subjects in each cohort.

	Cohort-1 Clinical High-Risk Included N=296		Cohort-1 Clinical High–Risk, Excluded N=81		Cohort-2 Clinical High-Risk Included N=592		Cohort-2 Clinical High–Risk, Excluded N=151	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Age	18.2	4.5	18.3	5.4	18.5	4.3	18.8	4.2
Parental education[1]	5.4	1.8	4.7	2.0	6.9	1.6	6.6	1.6
Scale of Psychosis-Risk Symptoms Total	38.9	14.5	36.5	12.6	38.2	12.3	37.1	11.9
2-Item Subscale	2.5	1.7	2.6	1.9	2.6	1.6	2.6	1.6
4-Item Subscale	3.3	2.0	3.2	2.1	3.8	2.0	3.6	2.0
	%	n	%	n	%	n	%	n
Ancestry, Caucasian Central/South America African Asian Multiracial Other	78.7% -- 9.1% 4.7% 5.1% 2.3%	233 -- 27 14 15 7	75.3% -- 8.6% 2.5% 8.6% 4.9%	61 -- 7 2 7 4	57.9% 4.1% 15.0% 7.3% 12.5% 3.2%	343 24 89 43 74 17	54.7% 6.0% 16.7% 6.7% 14.0% 1.7%	83 9 25 10 21 3
Sex, male	60%	178	72%	58	58%	54	53%	76
High risk syndrome (not mutually exclusive):								
attenuated psychotic symptoms	96%	284	94%	69	95%	562	94%	142
brief intermittent psychotic symptoms	3%	9	6.2%	4	3%	18	1%	2
genetic risk and functional deterioration	13%	38	12%	6	12%	71	9%	18

[1] In Cohort-1 included n=215 & excluded n=55; In Cohort-2 included n=585, excluded n=141, SES was significantly lower in persons with no follow-up visits to those with at least one follow-up visit, Cohort 1 p-value= 0.02, Cohort-2 p-value=.05

Table 2

Evaluation of discrimination in the derivation Cohort-1* and independent test Cohort-2*.

	2-Item Subscale				4-Item Subscale				Scale of Psychosis-Risk Symptoms Total			
	Cohort-1		Cohort-2		Cohort-1		Cohort-2		Cohort-1		Cohort-2	
	value	95% CI	value	95% CI	value	95% CI	value	95% CI	value	95% CI	value	95% CI
Hazard function[1]	2.07	1.63, 2.62	2.14	1.65, 2.78	3.09	2.27, 4.21	2.99	2.08, 4.29	2.21	1.66-2.94	1.84	1.31-2.58
Concordance Index[2]	2.07	0.66-0.79	0.68	0.63, 0.76	0.74	0.69, 0.80	0.71	0.64, 0.77	0.68	0.63-0.74	0.61	0.55-0.67

[1]To calculate the hazard function the scales were divided by the number of symptoms used to form the scale. The hazard represents the increase in risk associated with changing an average of one unit. The hazard function for the 4-Item vs. Scale of Psychosis-Risk Symptoms Total scale in Cohort-1 p=0.005, Cohort-2 p=0.001.

[2]At 2-years in Cohort 2 at 2-years in 80% randomly selected (with replacement) subject subsets the C-Index was higher for the 4-Item than for the 2-Item for 743 of 1000 trials, the C-Index was higher for the 4-Item subscale than the SOPS Total for 968 of 1000 trials, and the C-Index was higher for the 2-Item subscale than the SOPS Total for 883 of 1000 trials, p-value<E-55.